Computer Fraud and Security
ISSN (online): 1873-7056

REACH: Remote Evaluation and Analysis with Certificate-free signcryption for community-level Health in Diabetic Retinopathy Screening

Zhike Han[1,2], Hanyu Xiao[1,*], QingQing Zheng[3], Tin Shen[4], Chaoyang Hong[3], Huang Xu[1], Shuiguang Deng[2]

[1]Hangzhou City University, Hangzhou, 310015, China
[2]Zhejiang University, Hangzhou, 310027, China
[3]Department of Ophthalmology, Zhejiang Provincial People's Hospital, Hangzhou, 310014, China
[4]School of Medicine, Zhejiang University, Hangzhou, 310009, China
**Corresponding Author.*

Abstract:
Diabetic retinopathy (DR) is a common complication of diabetes that often does not show symptoms in the early stages, so timely screening and diagnosis are crucial. This study proposes a remote DR screening system called REACH. This system integrates a portable fundus imaging device, an artificial intelligence (AI) inference module, and an information management system to make the DR screening process more efficient. The REACH system uses a deep learning model to analyze and classify fundus images. The experimental results demonstrate that the system achieved excellent classification performance on the DDR, APTOS, and EyePACS datasets, with accuracy rates of 89.48%, 85.31%, and 85.85%, respectively, and Kappa coefficients of 0.9297, 0.9207, and 0.9043. Additionally, the REACH system achieved an initial screening accuracy of 88.89% in clinical trials conducted at Zhejiang Provincial People's Hospital and three affiliated primary care hospitals. These findings indicate that the REACH system has the potential to significantly enhance the coverage and efficiency of DR screening in regions with limited medical resources. In the data collection process, an unsigned encryption algorithm is integrated to unify the signing and encryption functions, making it suitable for resource-constrained remote areas. This approach ensures secure data transmission and protects patient privacy.

Keywords: diabetic retinopathy, deep learning, artificial intelligence, certificateless signcryption, computer-assisted diagnosis

INTRODUCTION

Diabetes is a widespread global disease that is rapidly increasing its prevalence. According to the latest 10th edition of the Global Diabetes Atlas released by the International Diabetes Federation, approximately 537 million adults worldwide had diabetes in 2021. This number is projected to increase to 643 million by 2030 and could reach 783 million by 2045[1]. Diabetic retinopathy (DR), a common complication of diabetes, is caused by prolonged hyperglycemia, which damages the small blood vessels in the retina and can lead to severe vision problems. Almost all patients with type 1 diabetes will develop retinopathy after 20 years, and even in patients with well-controlled type 2 diabetes, 70%–80% may develop retinopathy after 15 years[2]. This scenario indicates that hundreds of millions of diabetic patients worldwide are at risk of, or are currently suffering from, retinopathy.

Eye screening programs for diabetic patients have been implemented in China to mitigate the impact of DR[3]. Existing screening methods rely on professionals using desktop equipment to capture and transmit retinal images. These methods usually require patients to visit specialized medical facilities for multiple consultations and examinations. While these screening approaches effectively detect DR in its early stages, their complicated processes and reliance on specialized equipment increase the burden on patients, especially in remote areas where medical resources and qualified personnel are scarce, leading to low screening coverage.

The traditional methods for screening DR also have other significant limitations. First, they rely heavily on expensive equipment, such as desktop fundus cameras and specialized devices[4]. Second, trained ophthalmologists or technicians must operate and interpret the images, which can be challenging in resource-limited and remote areas[5]. Patients often must make multiple trips to healthcare facilities, which adds to their time and financial burden[5].

This study proposes the REACH system, a novel solution to streamline the DR screening process to address these challenges. The REACH system reduces the reliance on expensive equipment and specialized personnel by combining a portable fundus camera, Certificate-free signature encryption algorithm, artificial intelligence (AI)-powered preliminary screening, and mobile applications. This system enables screenings in community settings or even in patients' homes, reducing the necessity for multiple visits to healthcare facilities. Moreover, the REACH system is well-suited for resource-limited environments, improving screening coverage in remote areas. In experiments, the AI inference module of the REACH system achieved accuracy rates of 89.48%, 85.31%, and 85.85% on three publicly available datasets—DDR, APTOS, and EyePACS—along with Kappa coefficients of 0.9297, 0.9207, and 0.9043, respectively. In clinical trials conducted at Zhejiang Provincial People's Hospital and three affiliated community hospitals, 36 patients were screened, including 18 DR patients and 18 normal individuals, with REACH achieving a preliminary screening accuracy of 88.89%. These results demonstrate that the REACH system exhibits high accuracy across multiple public datasets and clinical trials, indicating its potential for broad applications in the future.

This study contributes the following:

- We develop a novel DR screening system called REACH, which integrates image acquisition, information management, and AI classification into a unified system. This integration enhances screening efficiency and accuracy.
- We introduce relationally weighted labels to improve the image classification accuracy. This scenario involves assigning weight labels based on the similarity or correlation between categories, thereby improving model prediction accuracy.
- The REACH system is designed explicitly for resource-limited environments, making conducting DR screening and management in remote areas more accessible. Incorporating a certificate-free signature encryption algorithm at the data collection end integrates the signing and encryption processes, making it suitable for resource-limited remote areas. This approach ensures secure data transmission and protects patient privacy.
- For the first time, we address the processing of low-quality fundus images by evaluating image quality for low-resolution images, achieving promising results in clinical tests.

RELATED WORKS

Substantial progress has been made in DR screening in recent years. Researchers have proposed various methods and systems ranging from traditional approaches based on lesion detection to more recent intelligent screening systems driven by deep learning[6]. Some of these systems have already been implemented, showing great potential for automated screening, early diagnosis, and enhancing healthcare efficiency[6].

A research team from King's College London and St Thomas' Hospital has developed an early tool for screening DR[7]. This tool uses image processing techniques to detect microaneurysms (MAs), hemorrhages (HM), and hard exudates (ED) and grades DR using color retinal images. Earlier versions were based on neural networks for detecting DR features, and Sinthanayothin et al. later proposed a series of improved algorithms[8]. These improvements include detecting the optic disc, macular fovea, blood vessels, and other abnormal features. Essential methods include local adaptive contrast enhancement, image intensity normalization, back-propagation neural networks for blood vessel detection, a recursive region-growing segmentation algorithm for exudate detection, and "trench operators" for automatically detecting MAs and hemorrhages. Based on these algorithms, Usher et al. developed an automated DR grading tool that involved image pre-processing, standard retinal component detection, candidate lesion detection and classification, and rule-based image grading[8]. In a screening of 1273 patients, this tool demonstrated clear clinical value, with a sensitivity of 95.1% and a specificity of 46.3% for DR detection, while ensuring no missed detections of vision-threatening DR [7].

Cunha-Vaz et al. developed the RetmarkerSR system for automated grading and screening of DR[9]. This system uses structural changes, such as the formation and disappearance of MAs, as early indicators of DR progression. It assesses DR progression by comparing fundus images taken at different time points. The grading process in RetmarkerSR is a two-step approach: First, MAs, and hemorrhages are detected to classify images as "with

disease" or "without the disease." Second, image registration techniques compare baseline images with current ones to assess structural changes. The system also integrates an image quality assessment algorithm proposed by Dias et al., which determines image suitability for analysis with a sensitivity of 96.67%[10]. Experimental results demonstrate that RetmarkerSR can indicate early DR activity through MAs turnover, showing higher sensitivity than methods relying solely on MAs counts. In large-scale screening tests, the system employed support vector machines (SVM) for "disease/no disease" classification with a sensitivity of 95.8%, highlighting its potential to reduce specialist workloads and screening costs.

The Tennessee Ophthalmic Telemedicine Network (OTN) was developed to provide an efficient, low-cost solution for population-based DR screening[11]. The system uses a web-based application that allows primary care clinics to upload patient data, enabling ophthalmologists to analyze images through a server-side application and automatically generate diagnostic reports. Critical functions of OTN include image quality assessment, automated image analysis, and diagnosis through content-based image retrieval (CBIR) technology to estimate DR status. The team also developed a vessel structure-based image quality assessment algorithm, demonstrating high classification and speed performance across multiple datasets[12]. Additionally, OTN employs several DR grading algorithms, such as the Radon cliff operator for detecting MAs and the Fisher linear discriminant classifier for distinguishing bright lesions. The CBIR method calculates the similarity between lesion features and the feature vectors stored in a database to retrieve lesion statuses quickly. Experimental results show that the OTN system achieves high sensitivity in DR screening, significantly reducing costs and relieving specialists' workload[13].

Abràmoff et al. developed a DR screening system. This system assesses image quality and detects retinal structures such as blood vessels, the optic disc, and the macula. The system subsequently detects potential lesions at pixel and cluster levels, using a fusion-based classification method to determine whether a patient has DR. The core algorithms include optic disc and macula localization, vessel segmentation, red lesion detection, and bright lesion detection (exudates, cotton-wool spots, and vitreous warts)[14]. Niemeijer et al. proposed an automated method based on image structure clustering and SVM classifiers for image quality assessment, which achieved an AUC of 0.9968 on 1000 screening images[15]. The system also integrates the Spencer–Frame method with a machine learning-based approach to detect red and bright lesions, classifying them using a k-nearest neighbors classifier and a linear discriminant classifier. In testing, the system achieved an AUC of 0.85 in the EyeCheck project with 10,000 screenings of 5692 patients. After incorporating information fusion, the AUC improved to 0.881 across 15,000 screenings. These results indicate that DR grading methods have improved. However, future work should focus on testing DR detection systems on more extensive and diverse diabetic populations.

The VisionQuest Biomedical team proposed an automated for detecting DR using an amplitude modulation-frequency modulation (AM-FM) approach. Unlike traditional "bottom-up" methods, this system adopts a "top-down" approach to detect retinal abnormalities[16]. It uses a modified AM-FM method for DR feature detection and employs partial least squares and SVM for DR classification. Instead of explicitly segmenting lesions, the system differentiates normal from abnormal images through texture modeling[17]. The core algorithms include multi-scale AM-FM feature extraction, principal component analysis, and k-means clustering for feature extraction and classification. Experimental results on the MESSIDOR dataset indicate that the system achieved an AUC of 0.89 for DR detection and 0.92 for vision-threatening DR. It demonstrates a sensitivity of 0.92 and a specificity of 1.00 for clinically significant macular edema (ME) [18].

A diagnostic system for grading DR and assessing the risk of ME was developed by the Centre for Mathematical Morphology (CMM) [19]. The system utilizes automated fundus image analysis algorithms developed over the past decade. These algorithms include optic disc and vessel detection, image quality assessment, and detecting MAs and exudates for grading DR and ME. A vascular tree-based method evaluates image quality based on vessel visibility. The system's core algorithms include optic disc localization, vessel detection using morphological techniques, and automatic detection of MAs and exudates. For DR grading, the main tasks involve detecting vessels, the optic disc, and the macula, along with measuring MAs and hemorrhages to assess DR severity. Exudate distribution in the macula is used to evaluate the risk of ME. Testing on the MESSIDOR dataset showed that the system achieved a sensitivity of 83.9% and a specificity of 72.7% for DR grading, a sensitivity of 72.8%, and a specificity of 70.8% for ME risk assessment[19].

Quellec et al. developed the TeleOphta system, an automated DR screening system based on 25,702 examination records from the OPHDIAT screening network[20]. The system enhances the PSM DR and ME grading systems' algorithms, particularly for detecting MAs and ME, and introduces data mining techniques for DR screening. Zhang et al. optimized MA and exudated detection methods using morphological sequence filters and Random Forest classifiers. They trained large annotated datasets to predict pathology probabilities in candidate regions. In validation tests on the e-ophtha EX dataset, the system achieved an AUC of 0.95 for exudate detection, with a sensitivity of 96% and a specificity of 89%. The TeleOphta framework also includes quality metrics and feature-based pathology scoring, combining lesion information, image quality, patient demographics, and diabetes-related data for detecting DR and other retinal conditions[21]. In tests on a subset of the e-ophtha dataset (500 records, 2 fundus images per eye), the system achieved a specificity of 68.0% and a sensitivity of 80.9%, aligning with results from a second expert reading[21].

- Many DR screening systems lack effective collaboration between academic institutions and healthcare organizations during development. This lack of cooperation may limit the systems' impact on widespread adoption and implementation, particularly in public health initiatives or national screening programs.
- The reliance of many systems on large, immobile imaging equipment is costly, limiting their use in resource-limited areas or for large-scale screening and presenting a significant challenge.
- Some DR screening systems face challenges with interoperability when integrating with existing healthcare systems, particularly electronic health records (EHR), which can affect their practical use in hospitals or clinics.
- Some systems also lack effective user interaction and feedback mechanisms, hindering their ability to incorporate feedback from doctors, nurses, and patients, which could optimize functionality. This limitation may reduce user engagement and the overall effectiveness of the system.
- Several systems struggle to adopt and integrate cutting-edge technologies, such as deep learning models and attention mechanisms, resulting in limitations when processing low-quality data.
- Many studies have used small and unevenly distributed datasets for training and testing, failing to adequately demonstrate the robustness and broad applicability of the algorithms in real-world clinical settings.
- Despite extensive testing on publicly available datasets, some studies lack validation in natural clinical environments, leaving the systems' effectiveness and stability unproven in practical screening scenarios.
- Some systems focus solely on a single screening task, failing to cover the broader healthcare ecosystem's needs, such as mobile tracking, expert advice, and integrated management information systems, creating gaps in functionality.

The existing DR screening systems have several limitations regarding classification accuracy, device convenience, dataset quality and scale, system integration, and scalability. Although these systems have improved in automated diagnosis and lesion detection, their limitations restrict their widespread application in natural clinical environments. This study introduces the REACH system to tackle these challenges, aiming to provide a comprehensive, efficient, and scalable solution for DR screening through a multi-layered system architecture, advanced AI inference algorithms, and efficient data management and user interaction mechanisms. The next section provides a detailed description of the REACH system's architecture and the implementation of its subsystems.

METHODS

Requirements and Architecture

A detailed description of the overall architecture of the REACH system and its various subsystems is provided. The system architecture comprises several subsystems, each responsible for specific tasks to conduct the DR screening process efficiently and reliably. The study starts with an overview of the system's overall design and its stakeholders in real-world applications, followed by an in-depth discussion of the functionality and

implementation of the individual subsystems as depicted in figure 1. These subsystems include the Fundus Data Acquisition Subsystem (FDAS), Screening Information Management Subsystem (SIMS), Data Storage Subsystem (DSS), Artificial Intelligence Inference Subsystem (AIIS), and the Image Quality Assessment Subsystem (IQAS).

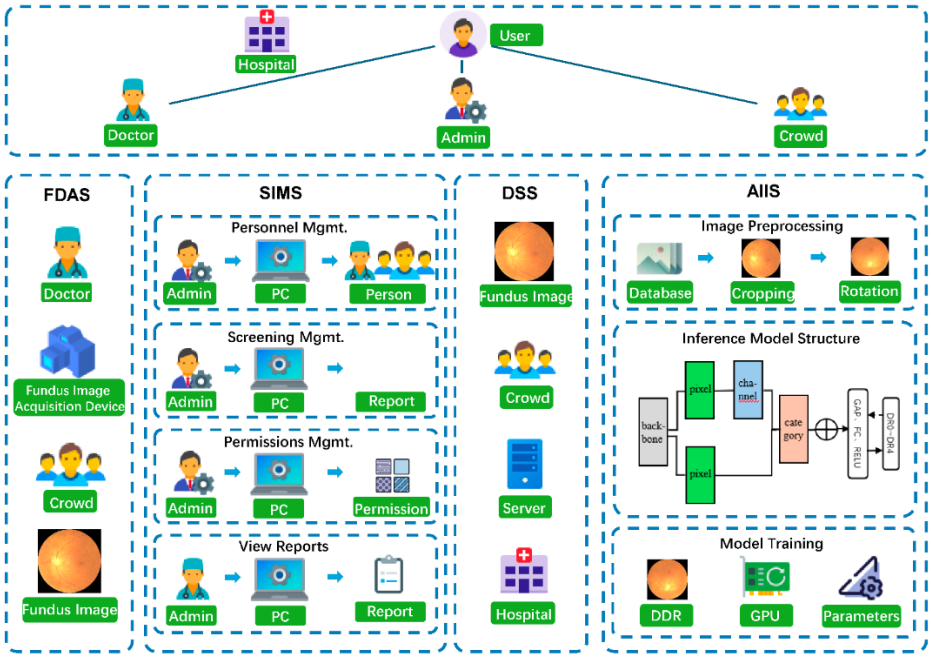


Figure 1. Architecture of the REACH system

The research involved developing a multi-layered system architecture to support complex data processing and user interaction. The system's inference module was implemented in Python using the PyTorch framework, chosen for its robust machine-learning libraries and flexible design. MySQL was selected as the database solution due to its stable and scalable data management capabilities. The front-end development utilized the VSCode editor and the Vue framework, which offered efficient and easy-to-manage user interface development through reactive programming and a component-based architecture. The back-end was developed in Java using the IntelliJ IDEA development environment and the Spring Cloud framework, which supported the construction of a microservice architecture due to its solid scalability and seamless communication between microservices. The system architecture ensures the use of advanced technologies and guarantees high development efficiency and system reliability.

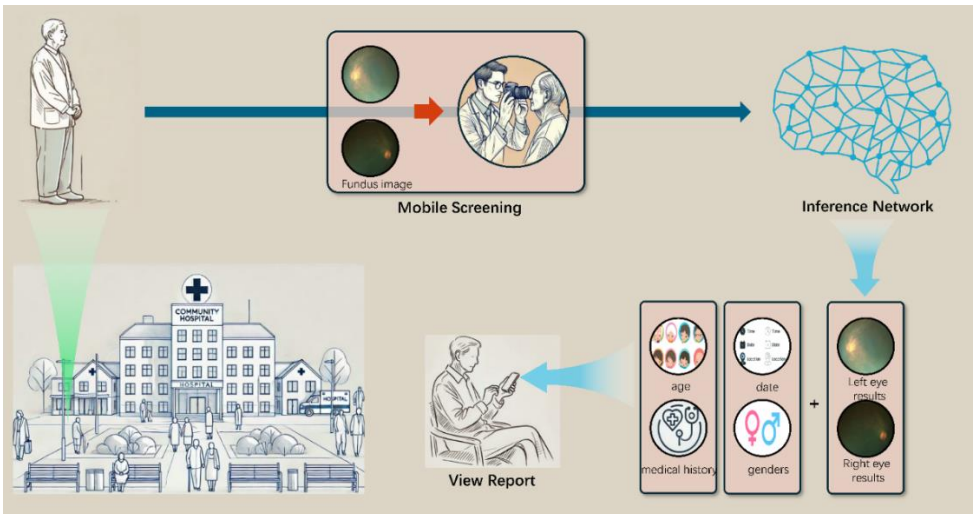


Figure 2. Community hospital screening scenarios

The REACH system uses a screening process, as depicted in Figure 2. First, patients' fundus images are taken using the FDAS at community hospitals, producing high-quality fundus images. Then, these images are transmitted over the network to the AI inference system, which automatically analyzes the images of both the left and right eyes, producing preliminary screening results. These results are combined with other patient information, such as age, medical history, gender, and examination date, to generate a comprehensive initial screening report. Finally, patients can access the screening report via mobile devices to monitor their eye health in real time. This process significantly improves the efficiency and accuracy of primary screenings at community hospitals, simplifies the patient screening experience, and brings convenience and intelligence to telemedicine services.

The key stakeholders involved in the design of the REACH system include the following categories:

- Doctors: Doctors are the primary users of the system. They use the REACH system to capture fundus images of patients and leverage the system's AI inference capabilities for preliminary diagnosis, significantly improving diagnostic efficiency and reducing misdiagnosis risk.
- Hospitals: Hospitals, the institutions where the system is deployed and used, efficiently manage patient screening data through the system. The system's low-cost design facilitates the implementation of primary healthcare facilities.
- Administrators: System administrators are responsible for maintaining the system, managing user access rights, and ensuring data security. They oversee the system's stable operation and safeguard the data within it.
- Patients: Patients benefit by receiving diagnostic results more quickly and with minimal impact from the screening process.

The REACH system comprises several shared subsystems, each responsible for a specific function within the overall system. These subsystems include:

- Fundus Data Acquisition Subsystem (FDAS): This subsystem captures patients' fundus images and transmits the data to other subsystems.
- Data Encryption Subsystem(DES): For data collection on mobile devices, especially in remote mountainous areas, a certificate-free signature encryption algorithm can be added at the data collection end to unify the signing and encryption processes.
- Screening Information Management Subsystem (SIMS): This subsystem manages all information related to the screening process, including patient data, screening records, access control, and report viewing.
- Data Storage Subsystem (DSS): The subsystem provides a secure and stable data storage solution, ensuring the integrity and safety of patient data.
- Artificial Intelligence Inference Subsystem (AIIS): This subsystem analyzes fundus images using deep learning models to generate DR grading results.
- Image Quality Assessment Subsystem (IQAS): This subsystem assesses the quality of fundus images to ensure that those submitted to the Artificial Intelligence Inference Subsystem (AIPS) subsystem meet the diagnostic standards.

By collaborating with these subsystems, the REACH system can provide comprehensive technical support for the DR screening process.

Fundus Data Acquisition Subsystem

The FDAS is designed to efficiently and accurately acquire images for DR screening. This subsystem includes a high-definition fundus camera, a data transmission module, and an intelligent terminal. It acquires high-definition images through optimized optical design and an automated focusing mechanism without pupil dilation. The system's core components include the housing, communication module, eyepiece lens, focusing lens, 1/4 wave plate, and CMOS sensor. The eyepiece lens system, consisting of three lenses combined with the dual 1/4 wave

plate configuration, optimizes the optical path and enhances imaging quality. Figures 3 depict the physical of the subsystem, respectively.

Figure 3. Physical image of the FDAS fundus image acquisition module

The system follows an integrated design concept, with the core components including a housing, communication module, eyepiece lens, focusing lens, the first 1/4 wave plate, and a CMOS sensor. The EIAM is equipped with a professional optical system arranged in the direction of light transmission with the eyepiece lens, focusing lens, the first 1/4 wave plate, and the CMOS sensor in sequence. The eyepiece lens comprises three lens elements, with a second 1/4 wave plate added between the first and second lenses. This layout optimizes the optical path, and the focusing lens is driven by a miniature stepper motor, enabling the capture of clear images without the need for pupil dilation.

The FDAS uses a multi-light source design, which includes adjustable-intensity white LEDs, low-power infrared light sources, and green LEDs. This design ensures that high-quality images can be acquired under various lighting conditions. The green LED is specifically used to guide fixation, enabling precise pupil alignment with the optical system and preventing common image distortion issues. The system employs a combination of linear and reflective polarizers to reduce light source interference, achieving effective optical isolation and ensuring that only scattered light signals from the fundus are captured during imaging.

The FDAS can transmit captured images to the intelligent terminal in real time through a high-speed USB data transmission interface. These images are then combined with the AIPS subsystem for rapid analysis and diagnostic feedback. This design simplifies the operational workflow and significantly enhances the efficiency of image acquisition and analysis. Experimental results indicate that the system maintains high image quality within a refractive range of −10D to +10D, and the full-field distortion is controlled within 3.5%, meeting the requirements for clinical diagnosis.

Data Encryption Subsystem

The data encryption subsystem plays a critical role in the Reach system, responsible for encrypting personal privacy information of the screened population. Existing certificate-free signcryption schemes often lack sufficient security, failing to fully withstand attacks from malicious Key Generation Centers (KGC) and malicious users. The scheme used in the data encryption subsystem builds upon traditional certificate-free encryption by integrating both signing and encryption functions, ensuring message confidentiality and integrity in a single operation.

Setup During initialization, the Key Generation Center (KGC) selects two cyclic groups G_1 and G_2, both of prime order p, and chooses a generator g. A bilinear mapping $e: G1 \times G1 \rightarrow G2$ is defined to support subsequent signcryption operations. Additionally, three collision-resistant one-way hash functions, H_1, H_2, and H_3, are selected to ensure uniqueness and collision resistance of the hashing process. The KGC also chooses random vectors $U = (u_i)$ and $M = (m_i)$, along with a random string K, which serve to enhance the scheme's attack resistance and increase the complexity of key generation.

Partial Private Key Generation In this step, the KGC generates an identity-linked partial private key psk_u based on the user's identity (e.g., email or phone number) and the master private key. For each user identity u, the KGC maps identity information with the master private key to create a unique partial private key. The randomly selected vectors U, M, and string K contribute to the generation of the partial private key, making the process more complex and ensuring the uniqueness and unpredictability of each user's partial private key, thus enhancing the key's security. This partial private key will be used by the user to generate a complete private key.

User Key Generation In this stage, the user selects a random secret value X_u and combines it with the partial private key to generate their public and private keys. This key generation method ensures the user's autonomy over the key, eliminating dependence on the KGC and addressing the key escrow problem. The user's public key is generated by combining the user-chosen random value with the partial private key provided by the KGC, ensuring security and independence in the key generation process, and reducing the risk of identity information leakage.

Private Key Extraction The user obtains the final complete private key sk_u by combining their randomly selected secret value X_u with the partial private key psk_u from the KGC. This step is essential in certificate-free encryption systems, ensuring that the partial private keys generated independently by the user and the KGC form a unique complete private key, meeting the design requirements of a certificate-free system.

Signcryption In the signcryption step, sender A uses their private key to both encrypt and sign the message M. Specifically, the sender first selects a random number r and uses recipient B 's public key to encrypt the message, generating an encryption key $\sigma_1 = H_1\big(e(PK_B, r)\big) \oplus M$, which is then used to encrypt M. Next, the sender generates signature information, including a combination of the random number and hash values, to ensure message integrity. Finally, the sender sends the ciphertext $\sigma = (\sigma 1, \sigma 2, \sigma 3, \sigma 4, \sigma 5)$ to the recipient, who can verify the authenticity of the message using this ciphertext.

Unsigncryption Upon receiving the ciphertext σ, recipient B uses their complete private key to decrypt and verify the signature, ensuring both message integrity and authenticity. The recipient first calculates the hash verification value related to the ciphertext and checks the validity of the sender's signature. If the verification is successful, the recipient can decrypt and retrieve the original message $M = \sigma_1 \oplus H_1\big(e(PK_B, r)\big)$; otherwise, decryption fails. As this scheme supports publicly verifiable ciphertext, the recipient can verify the source and authenticity of the ciphertext without decrypting it, making this scheme suitable for applications requiring verification of information authenticity, such as spam filtering and firewalls.

The data encryption subsystem achieves higher efficiency during signcryption and unsigncryption phases, requiring fewer public parameters and keys, significantly reducing computational and memory overhead.

Screening Information Management Subsystem

SIMS plays a critical role in the REACH system as it manages and coordinates the flow of information throughout the DR screening process. SIMS consists of four submodules: the Personnel Information Management Submodule, the Screening Information Management Submodule, the Permission Management Submodule, and the Report Viewing Submodule. These submodules collaborate to ensure efficient information collection, processing, storage, and presentation. Figure 4 shows the main interfaces of the REACH system on mobile and desktop platforms.

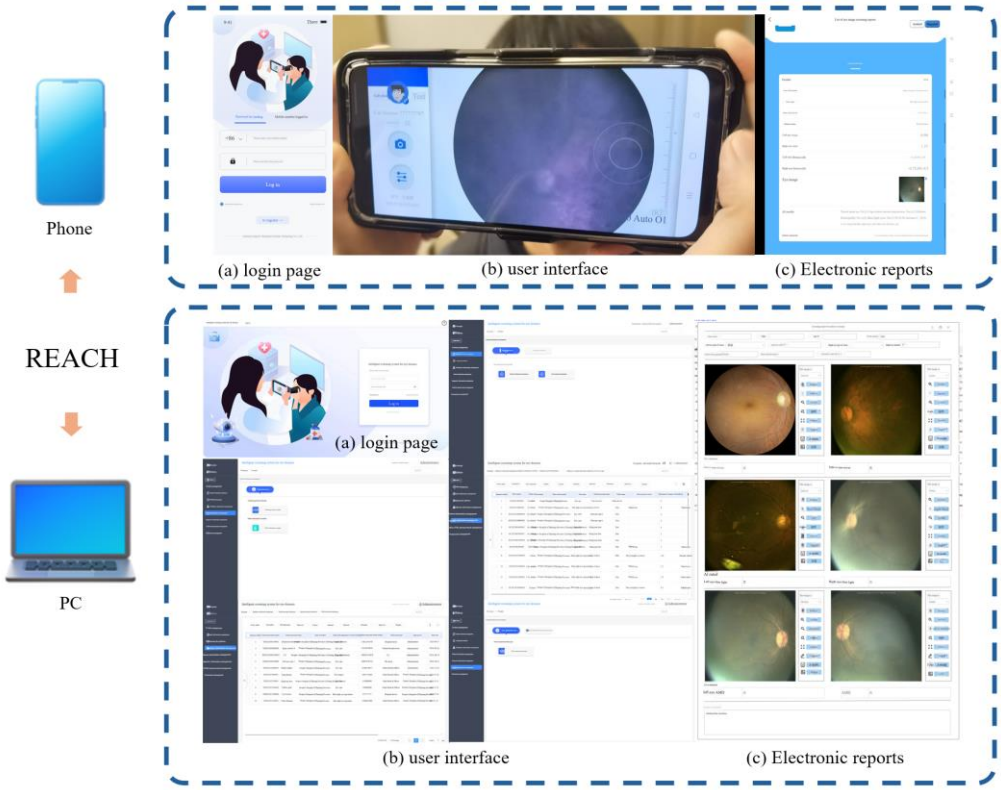


Figure 4. Main interfaces of the REACH system on mobile and desktop platforms

Personnel information management submodule

The Personnel Information Management Submodule collects and manages all personnel-related data during the screening process. It gathers data from multiple sources, including fundus images captured by the FDAS subsystem, healthcare workers' operational records, and patients' personal information. This data is automatically collected through integration with the FDAS subsystem and stored in a centralized database. Data processing involves cleaning, structured storage, and optimized handling based on query and retrieval needs. Ultimately, this data supports other submodules within the system, ensuring an efficient and accurate data management process throughout screening.

Screening information management submodule

The Screening Information Management Submodule manages essential core data related to DR screening, including screening results, screening history, and metrics. This submodule works closely with the AIPS subsystem to receive and process screening results, which are then stored in the database for healthcare professionals to access and analyze. The submodule maintains coordinated data management through predefined rules and procedures, supporting real-time data analysis and report generation.

Permission management submodule

The primary task of the Permission Management Submodule is to maintain data security within the system. This submodule controls user access rights using a Role-Based Access Control mechanism, which sets permission levels for different user roles to ensure that only authorized personnel can access sensitive data. Working alongside the system's security framework, the Permission Management Submodule provides flexible and highly secure permission management, preventing unauthorized access and potential data breaches.

Report viewing submodule

The Report Viewing Submodule offers an intuitive interface for healthcare professionals to access and analyze screening reports generated by AI inference. It receives processed image analysis data from the AIPS subsystem and presents it visually, showing screening results, historical trends, and detailed patient reports. The submodule

also enables healthcare professionals to review and adjust AI-generated results, enhancing the accuracy and reliability of the final diagnosis.

Data Storage Subsystem

The DSS is a critical component of the REACH system architecture. It ensures that the data generated during the DR screening process is securely and reliably stored and managed. The DSS uses advanced data management technologies to provide a highly secure environment for storing patient retinal images and related screening information.

During data storage and management, the DSS implements stringent data protection measures to meet compliance requirements for medical data. It supports data encryption, access control, and regular backups, ensuring the security and privacy of patient information throughout storage and processing. Additionally, the DSS leverages cloud storage technology to provide long-term data retention and flexible access, allowing data to be quickly and securely retrieved and analyzed as needed.

The DSS is designed for seamless integration with other subsystems, particularly regarding data flow and processing. Fundus images captured by the FDAS subsystem are directly uploaded to the DSS via predefined interfaces for storage. This stored data is then transmitted to the AIPS for further analysis and diagnosis. This tight integration ensures the automation and efficiency of the entire workflow, from image acquisition to the generation of analysis results.

By implementing these features, the DSS guarantees data security and reliability within the REACH system and enhances its overall synergy, supporting its rapid deployment and scalability.

Artificial Intelligence Inference Subsystem

Model training

After designing the network architecture design, the AIIS subsystem was trained using the pre-processed DDR dataset to make the model effectively learn and accurately predict the severity of DR. The model was deployed on a high-performance GPU server (model: GeForce RTX 2080Ti), with a learning rate of 0.0001 applied for 50 training epochs. This learning rate was validated through multiple experiments to strike a good balance between convergence speed and model stability. Figure 5 illustrates a machine learning pipeline for medical image analysis using DDR, APTOS, and EYE-PACS datasets. The process begins with data cleaning and preprocessing, followed by splitting the data into training, testing, and validation sets, with the validation set incorporating a clinical dataset for real-world applicability.

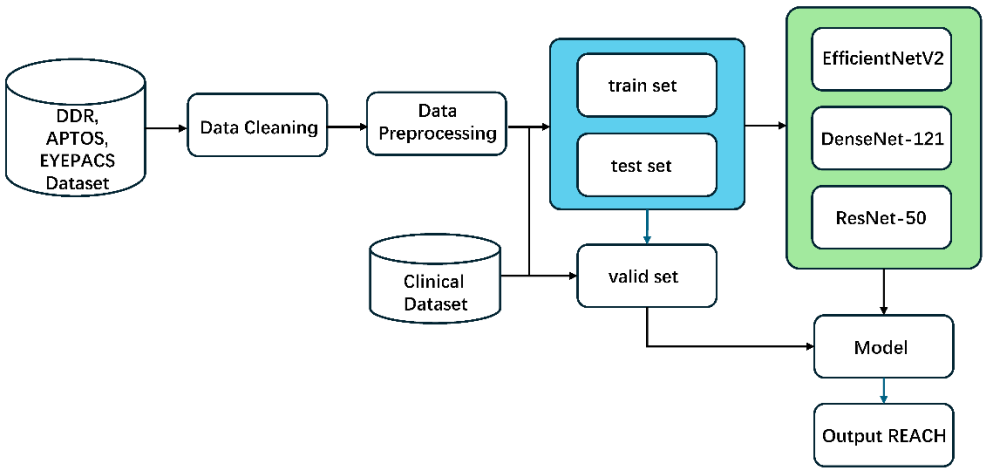


Figure 5. Training Process

The model's performance is dynamically monitored throughout the training process using key metrics, such as accuracy, loss, and the Kappa coefficient. An early stopping strategy was implemented to prevent overfitting, which stops training as soon as a downward trend in validation accuracy is detected. Figure 6 shows the training

curves for the model with and without using relationally weighted labels, displaying the trends in loss, accuracy, and Kappa coefficient.

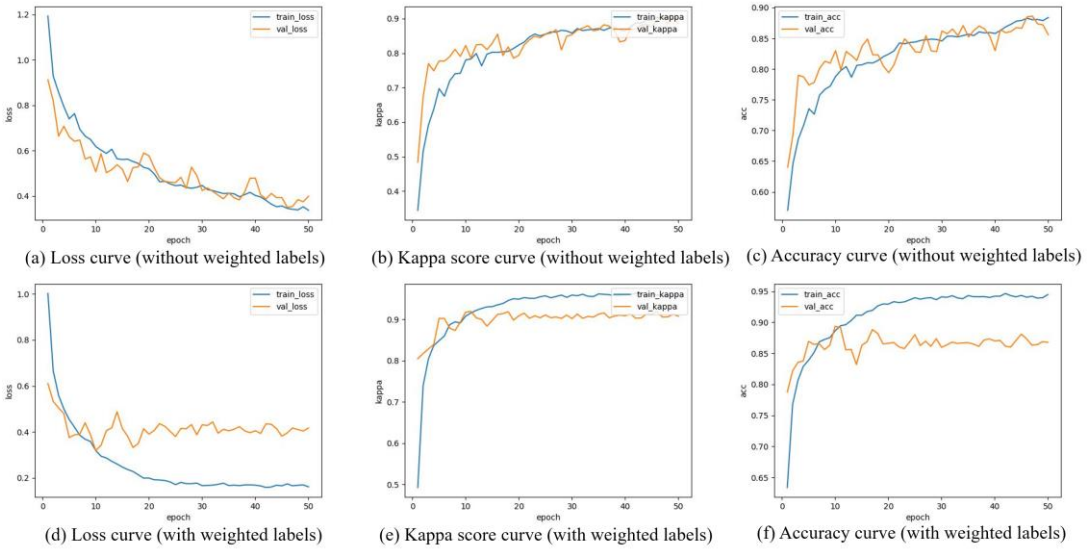


Figure 6. Training curves. the dataset used is the DDR dataset. Subfigures (a)-(c) correspond to training without weighted labels, while (d)-(f) correspond to training with weighted labels. The loss curve, kappa score curve, and accuracy curve are shown from left to right, respectively.

Introduction to inference networks

The AIIS is a core component of the REACH system. Its primary purpose is to predict the severity of DR by analyzing fundus images. AIIS works closely with the high-resolution FDAS, receiving image data from FDAS and inputting it into a deep learning model for accurate predictions. To achieve high-precision DR grading, AIIS employs a multi-module deep learning network architecture comprising four key components: the Backbone, the self-attention mechanism, module arrangement, and relationally weighted labels, as depicted in Figure 7.

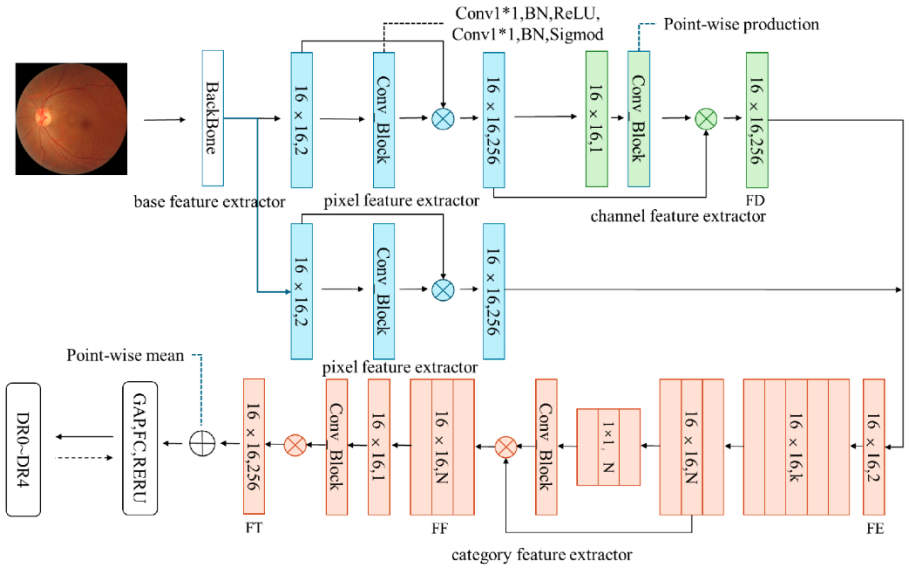


Figure 7. Relationship weight network structure

(a) Backbone: The AIPS network's infrastructure consists of three backbone networks: EfficientNetV2-s, DenseNet-121, and ResNet-50 [22-24]. Together, these networks are responsible for extracting the initial global features from the input fundus images:

- EfficientNetV2-s optimizes the network's width, depth, and resolution through a compound scaling approach, allowing the model to extract high-quality features while maintaining computational efficiency[22].
- DenseNet-121 facilitates gradient flow and alleviates the vanishing gradient problem through dense connections, providing a reliable foundation for subsequent feature extraction[23].
- ResNet-50 enhances information flow and strengthens model robustness through residual connections, allowing the network to excel at handling complex features[24]. These backbones were chosen based on their outstanding performance in various visual tasks[25], particularly their efficiency and accuracy in feature extraction and classification.

(b) Self-Attention Mechanism: The AIPS network incorporates several self-attention modules, including channel attention, spatial attention, and category attention modules, to enhance the model's feature-capturing ability. These modules selectively amplify the model's response to critical features using different attention mechanisms:

- Channel Attention Module: This module focuses on extracting features between image channels, treating feature maps of different channels as distinct detectors representing meaningful information in fundus images. It obtains the channel weight distribution through pixel-wise global average pooling and the conv_block method, which consists of a 1x1 convolutional layer, ReLU activation, batch normalization, and a sigmoid function. The mathematical expression for this section is presented as follows:

$$F_C = CB\big(GAP_S(F_B)\big) \otimes F_B \tag{1}$$

- Spatial Attention Module: This module targets feature extraction in the pixel dimension. Different pixels on the feature map represent significant locations in the fundus image. It obtains the pixel weight distributions from the global average pooling of the channel dimension and conv_block processing. The mathematical expression for this section is presented as follows:

$$F_S = CB\big(GAP_C(F_B)\big) \otimes F_B \tag{2}$$

- Category Attention Module: This module assigns various channels to each category based on weight, enabling the model to understand and extract features from multiple perspectives. The category channel pooling transforms channel-dimension features into category-dimension features. It continues by performing global average pooling and conv_block processing on the pixel dimension, focusing exclusively on feature learning within the category dimension. It concludes with the global average pooling in the channel dimension to determine the final category weight distribution. The calculation formula for the category extraction process is as follows:

$$K = \sum_{i=1}^{N} k_i \tag{3}$$

$$F_N = GMP_K\big(Conv_K(F_B)\big) \tag{4}$$

$$F_T = CB\big(GAP_C(F_N)\big) \otimes F_B \tag{5}$$

(c) Module Arrangement:The network's self-attention modules are arranged to allow the model to extract features. First, the channel attention module selectively strengthens important channel features to capture critical global information. Next, the spatial attention module focuses on the most distinctive spatial locations to improve computational efficiency and reduce interference from irrelevant information. Finally, the category attention module enhances the model's classification ability by assigning weights to different categories. This modular arrangement significantly improves the model's overall performance, increasing accuracy and robustness in the DR grading task.

(d) Relationally Weighted Labels: AIPS introduced the relationally weighted labels method[26] to optimize traditional label encoding approaches during model training. In DR grading tasks, there are significant distance relationships between categories. Conventional one-hot encoding[27] and label encoding[28] often overlook these relationships. By incorporating label distance weights, the relationally weighted labels method rationalizes the distance between categories, leading to improved calculation loss. This method offers notable advantages in

handling data imbalance, enhancing the model's classification accuracy and generalization capability. In this paper, the weight $weigh_{ij}$ is computed using the following formula:

$$\text{weigh}_{ij} = \left(1 - \frac{|i-j|}{N-1}\right)^2 \quad (6)$$

where i and j denote the actual and predicted labels, respectively, N is the total number of classes, and $| i - j |$ represents the distance between the two labels.

Image Quality Assessment Subsystem

Color fundus images are widely used to screen and diagnose fundus diseases such as DR, age-related macular degeneration, and glaucoma. However, the accuracy of these computer-aided diagnostic (CAD) systems heavily relies on the quality of the input images. Low-quality images, which are blurred, underexposed, or overexposed, can significantly impact the diagnostic accuracy of both physicians and CAD systems, reducing the reliability and effectiveness of the diagnosis.

This system incorporates an Image Quality Assessment Module (IQAM) to ensure the input images meet diagnostic quality standards. The IQAM uses a convolutional neural network (CNN) to classify the quality of fundus images. We chose the classic AlexNet architecture for image quality assessment due to its strong performance in image classification tasks. As a result, AlexNet can effectively identify and classify low-quality images, providing reliable input data for subsequent automated screening.

In practice, the IQAM module first passes the input fundus images through the pre-trained AlexNet model, which performs feature extraction and classifies the images based on their quality. This process allows the system to filter out images that do not meet diagnostic criteria, such as those degraded by blurring, insufficient lighting, or media opacity. The filtered, high-quality images are then used in subsequent disease screening workflows.

RESULTS

Dataset

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Laoreet id donec ultrices tincidunt arcu. Sollicitudin aliquam ultrices sagittis orci a scelerisque. Sit amet aliquam id diam maecenas ultricies mi. Proin fermentum leo vel orci porta non. Ornare arcu dui vivamus arcu. Lorem ipsum dolor sit amet consectetur. Cras fermentum odio eu feugiat pretium nibh ipsum. Sapien nec sagittis aliquam malesuada bibendum arcu vitae elementum curabitur. Rhoncus est pellentesque elit ullamcorper dignissim cras tincidunt lobortis feugiat. Venenatis urna cursus eget nunc scelerisque viverra mauris in. Diam volutpat commodo sed egestas egestas fringilla phasellus faucibus. Sit amet volutpat consequat mauris nunc congue nisi vitae. Tincidunt ornare massa eget egestas purus viverra accumsan in nisl. Semper quis lectus nulla at volutpat diam ut. Lobortis feugiat vivamus at augue eget arcu dictum varius duis. Vel facilisis volutpat est velit egestas dui id ornare arcu.

This study performed ablation experiments on the DDR dataset[29] and further validated additional DR datasets. The statistical information for these datasets is provided in the Table 1 below:

Table 1. Data volume statistics for standard public DR datasets

DataSet	DR0	DR1	DR2	DR3	DR4	SUM
DDR	6266	630	4477	236	913	12522
APTOS	1805	370	999	193	295	3662
EyePACS	65342	6205	13153	65342	1914	88701

The DDR dataset comprises over 1.6 million fundus images from more than 400 clinical hospitals across 26 provinces in China. These images underwent professional screening, annotation by ophthalmologists, and modeling evaluation by computer experts. We used 12,522 gradable images. The provided training, validation, and test sets contained 6260, 2503, and 3759 fundus images, respectively, with a DR0 to DR4 ratio of approximately 50:5:35:2:7. No modification was made on the original dataset splits, and we directly adopted the official divisions.

Unless otherwise stated, our experiments were primarily conducted on the DDR dataset. Models were trained on the provided training set, with the model achieving the highest Kappa score on the validation set selected for testing and result generation on the test set.

The APTOS dataset[30] was sourced from a Kaggle competition consisting of 3662 images collected by doctors in an Indian hospital using various devices. Due to the limitations of the capturing equipment and environment, most APTOS images lack complete eye information, often missing details from the upper and lower edges of the eye. Additionally, some images and labels contain noise. We randomly selected 70% of the data from each category for the training set for experiments on the APTOS dataset, with the remaining data used as the validation set to produce the final model results. The number of images in each category for the training and validation sets was 1805: 370: 999: 193: 295 and 541: 111: 299: 57: 88, respectively, with category proportions of 49: 10: 27: 5: 8.

The EyePACS dataset contains 35, 129 diabetic retinal fundus images categorized into five classes from DR0 to DR4. These images were captured by technicians from the Aravind Eye Hospital in India, who traveled to medically underserved rural areas and were later examined, classified, and labeled by experienced ophthalmologists. The image resolutions in the EyePACS dataset range from 433*289 pixels to 5,184*3,456 pixels, and some images exhibit defects such as blur, underexposure, or overexposure.

For experiments on the EyePACS dataset, we used the same data division method as for APTOS and applied random shuffling. After the split, the number of images in each category for the training and validation sets was 25809:2443:5292:873:873 and 8130:720:1579:237:240, respectively, with category proportions of 70:10:20:5:5.

Data Pre-processing

Data pre-processing is a critical step in this project, as the quantity and quality of data significantly impact the model's training results. Due to the unique attributes of medical data, we face data quantity and quality challenges, making data pre-processing a crucial aspect that requires focused attention and resolution.

Data pre-processing primarily involves two aspects. The first is dealing with poor-quality and indistinguishable images caused by objective factors. Such data should be excluded as they cannot provide helpful information. The second aspect involves handling data with annotation errors caused by subjective factors. These data should be retained and re-annotated. This can be achieved using a primary neural network, such as ResNet, to classify the raw data, identify the frequently misclassified portions, and re-annotate these by professional doctors. Additionally, poor-quality images in the dataset should be removed to minimize interference in model training and ensure that the focus of follow-up examinations is more targeted.

By examining the fundus images in significant DR datasets, we observe some common characteristics, such as varying image sizes and degrees of eyeball completeness and positioning. Ideally, the fundus images should be complete, precise, and centrally positioned within a circular frame. However, the completeness and position of the eyeball and the background color of the images are not factors that determine DR grading. Instead, they introduce irrelevant variables that can interfere with the model's extraction of significant features. Therefore, it is crucial to minimize these unnecessary variations as much as possible.

Evaluation Metrics

This study focuses on the multi-classification task for DR grading. We use accuracy and quadratic weighted kappa scores as evaluation metrics to demonstrate the model's performance. The drawback of relying solely on accuracy is that when the sample proportions are very imbalanced, the category with the more significant proportion can dramatically influence the accuracy.

Quadratic weighted kappa assigns different scores to different degrees of classification errors, aligning with our standards for assessing the DR severity. Therefore, it provides a satisfactory reflection of the model's performance than accuracy. When these two standards conflict, we prioritize the quadratic weighted kappa score as the primary reference standard.

For the five-class classification task, we calculate the accuracy and quadratic weighted kappa scores using the confusion matrix M. Let N represent the number of classes, and let M_{ij} denote the value at position i,j in the

confusion matrix, indicating the number of instances with accurate label iii predicted as class jjj by the model. The model's prediction is correct; otherwise, it is incorrect when i = j. Therefore, we can derive the formulas for calculating accuracy and quadratic weighted kappa score as follows:

$$\text{accuracy} = \frac{\sum_{i=0}^{N-1} M_{ii}}{\sum_{i,j=0}^{N-1} M_{ij}} \tag{7}$$

$$\text{kappa} = 1 - \frac{\sum_{i,j=0}^{N-1} W_{ij} \times M_{ij}}{\sum_{i,j=0}^{N-1} W_{ij} \times E_{ij}} \tag{8}$$

where:

$$W_{ij} = \left(\frac{i-j}{N-1}\right)^2 \tag{9}$$

$$E_{ij} = \frac{\left(\sum_{i=0}^{N-1} M_{ij}\right) \times \left(\sum_{j=0}^{N-1} M_{ij}\right)}{\sum_{i,j=0}^{N-1} M_{ij}} \tag{10}$$

Ablation Experiments

We initially conducted quantitative experiments to explore the impact of using three types of distance-based relational labels on the results of a hierarchically weighted network with category weight assignment. We analyzed the effect of applying relationally weighted labels on classification performance to validate the broad applicability of the proposed method. Finally, qualitative results are presented.

Experiment on selecting distance-relational label parameters

We compared the effects of three distance-relational labels: weightA, weightB, and weightC. The experimental results are shown in Table 2. Among the three types, weightC performed relatively better, with the accuracy and Kappa score increasing by 0.24% and 1.24%, respectively, when the value of *k* was uniformly set to 5.

The experimental results indicate that although weightA and weightB consider the relative distances between labels, their label differentiation ability is not as strong as that of one-hot encoding, leading to suboptimal grading performance. In contrast, weightC maintains a strong differentiation ability and accounts for the relative distances between labels, leading to improved prediction performance.

Table 2. Results of experiments with weighted labels for different relationships

K	Method	Acc	Kappa	δ Acc/%	δ Kappa/%
k=[5,5,5,5,5]	Baseline	0.8359	0.8553	-	-
	weightA	0.7984	0.8510	-3.75	-0.43
	weightB	0.8621	0.8569	-0.98	↑ 0.16
	weightC	**0.8383**	**0.8677**	**↑ 0.24**	**↑ 1.24**

Experiment on distance-relational label variability

We investigated the impact of using or not using relationally weighted labels with various backbone networks. Table 3 presents the experimental results. The results demonstrate that, regardless of the backbone network used, applying relationally weighted labels leads to better performance than not using weighted labels.

Table 3. Results of experiments on whether to use relationally weighted labels

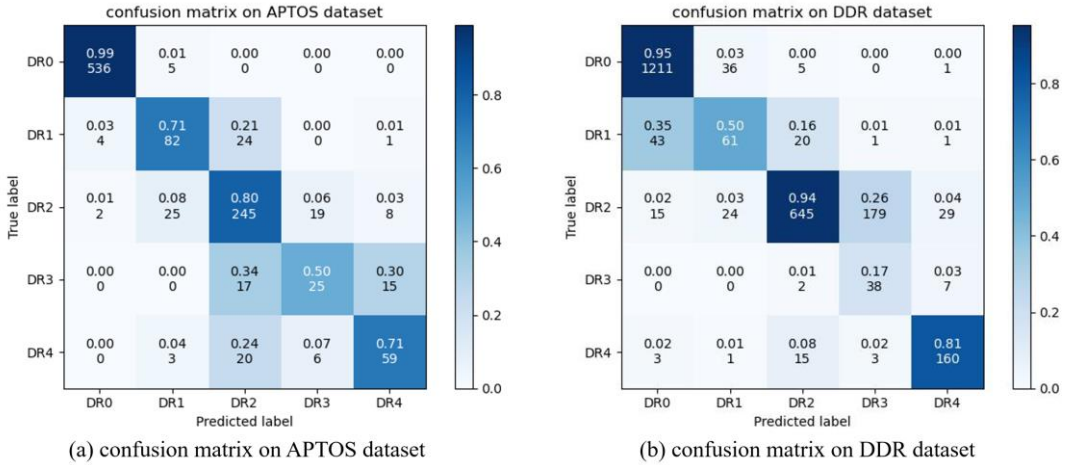
	Baseline	Acc	Kappa	Loss
No relational weighted labels	DenseNet121	0.8516	0.8605	0.4788
	ResNet50	0.8408	0.8462	0.4255
	EfficientNetV2-s	0.8848	0.8955	0.3501
relational weighted labels	DenseNet121	0.8924	0.9138	0.3444
	ResNet50	0.8896	0.9198	0.3753
	EfficientNetV2-s	**0.8948**	**0.9297**	**0.3548**

Comparison with Latest Results

To validate the effectiveness of our methodology, we compared the latest research results on three datasets, as shown in Table 4. For the DDR dataset, the officially provided training set was used for model training, and the validation set was used to test the final results. For the APTOS dataset, 70% of the data is randomly selected for model training, and the remaining portion is used to test the final results. The experiments used EfficientNetV2-s as the essential feature extractor for hierarchical, weighted networks. The confusion matrices shown in Figure 8 provide a detailed comparison of this method's performance on the APTOS dataset (a) and the DDR dataset (b). On the APTOS dataset, the confusion matrix shows high classification accuracy for DR0 and DR4, demonstrating the model's ability to distinguish between no DR and proliferative DR cases. However, some misclassification occurs between adjacent severity levels, particularly DR3 and DR4, due to the subtle visual differences in early stages. A similar trend is observed on the DDR dataset, where the model achieves robust performance for extreme cases (DR0 and DR4) but shows higher confusion between intermediate levels, such as DR2 and DR3. Figure 9 presents qualitative results for fundus images across different DR severity levels (DR1-DR4). Columns (a)-(d) represent images from different severity levels, while rows depict the original fundus images, activation maps obtained using only the DenseNet121 network, and heat maps generated by the hierarchically weighted network. In contrast, the heat maps from the hierarchically weighted network (third row) provide more concentrated and accurate localization of pathological regions, highlighting microaneurysms, exudates, and hemorrhages with greater clarity.

Table 4. Results on the DDR dataset, APTOS dataset, and EyePACS dataset compared to the latest research results

Dataset	Method	Acc	Kappa
DDR	DeepMT-DR	0.8360	0.8020
	CABnet	0.8569	0.8794
	CNN512	0.8860	0.8963
	Ours	**0.8948**	**0.9297**
APTOS	MSA-Net	0.8460	0.8960
	Mxception	0.8600	0.9000
	CNN512	0.8410	0.9062
	Ours	**0.8531**	**0.9207**
EyePACS	**Ours**	**0.8585**	**0.9043**



(a) confusion matrix on APTOS dataset (b) confusion matrix on DDR dataset

Figure 8. Confusion matrix of this paper's method on the APTOS (a) and DDR dataset (b)

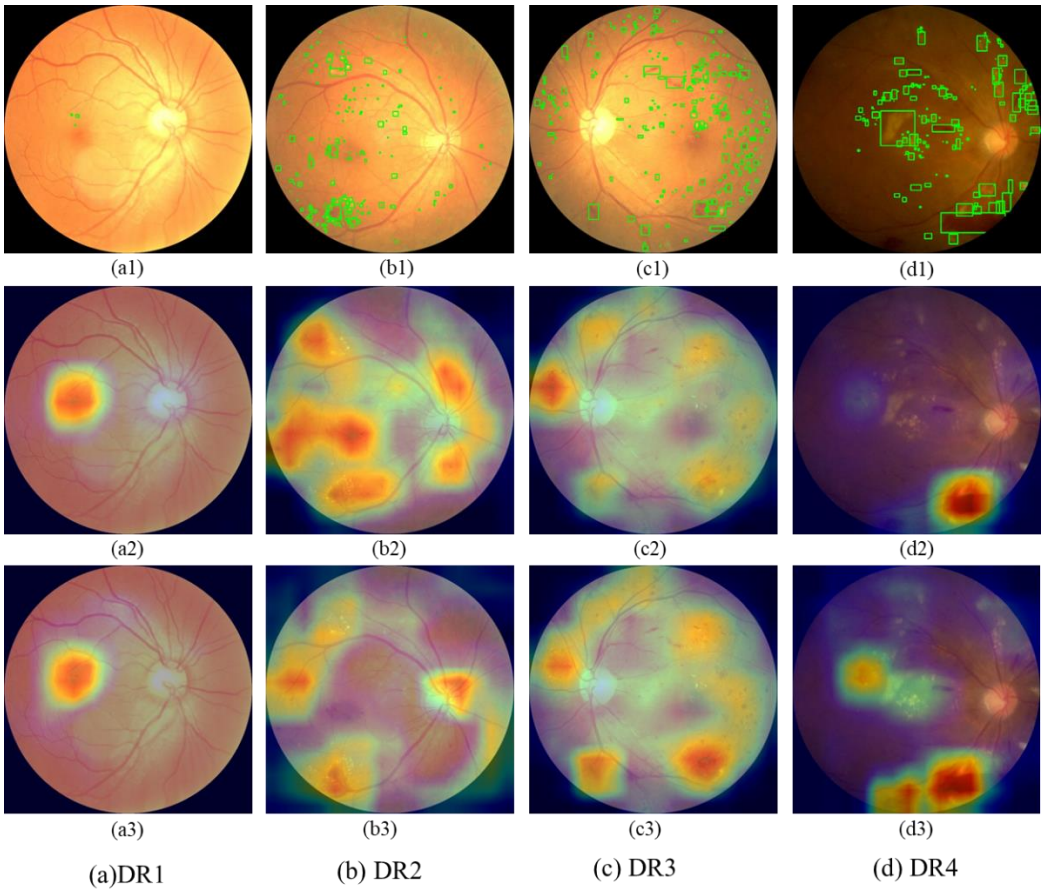


Figure 9. Qualitative results. Columns (a)-(d) correspond to DR1, DR2, DR3, and DR4, respectively. Rows represent the original images, the activation maps obtained using only the DenseNet121 network, and the heat maps obtained using the hierarchically weighted network (from 1-3).

Clinical Test Results

Clinical dataset

We created a clinical test dataset to assess the model's performance in a real-world clinical setting. The dataset consists of 36 anonymized RGB color fundus images and was developed in strict compliance with the World Medical Association's Declaration of Helsinki. All participants signed informed consent forms. Patients were recruited from the reporting and participating hospitals and consented to use the fundus imaging equipment for the study. All images were captured by the same physician on the same day to ensure consistency and reliability.

The clinical dataset comprises 18 images from DR patients and 18 from non-DR patients. Two expert doctors from Zhejiang Provincial People's Hospital labeled all images. A third expert reviewed and confirmed the final labels to ensure accuracy and consistency in labeling discrepancies. The images were graded on a scale from DR0 to DR4. Figure 10 compares the clinical dataset with the public dataset used during training, and Table 5 summarizes the distribution of images across the different DR levels.

Dataset characteristics

The clinical test dataset differs significantly from the public datasets used for image quality and characteristics training. First, the lighting conditions in the clinical dataset are suboptimal, with some images suffering from insufficient lighting. Additionally, the resolution of the clinical images is relatively low, and some images contain extraneous characters at the top, which may interfere with the model's evaluation.

Another notable difference is the lighting source: the training dataset used xenon lamps. At the same time, the clinical images were captured using white light illumination, leading to color discrepancies between the two datasets. These differences may affect the model's ability to recognize image features, potentially impacting its

performance on the clinical dataset. As a result, while the model performed well on the public dataset, the differences above may cause the clinical test results to be slightly lower than expected.

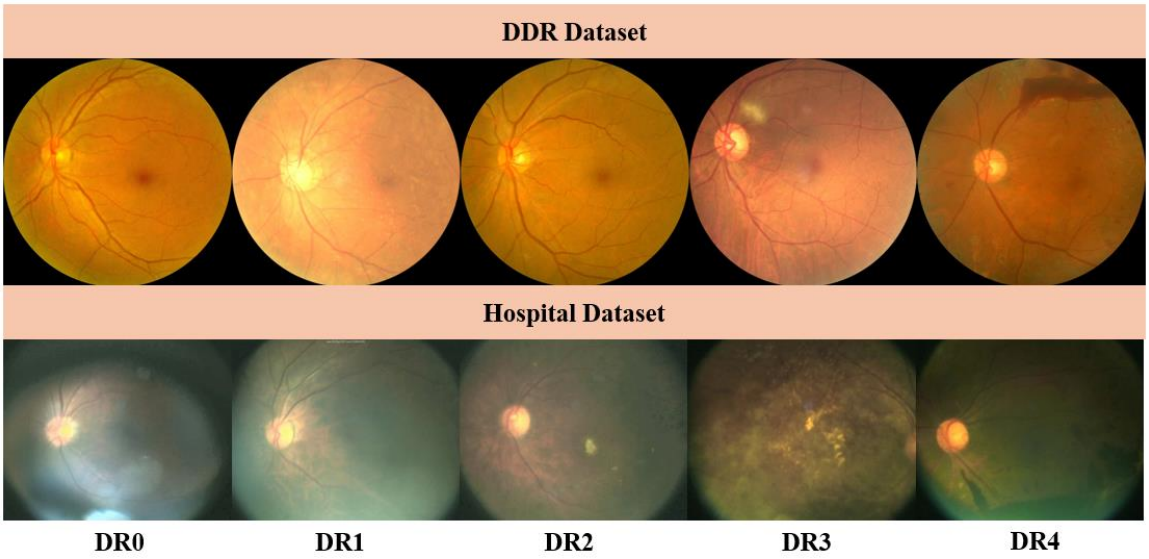


Figure 10. Hospital dataset vs. DDR dataset

Test results and analyses

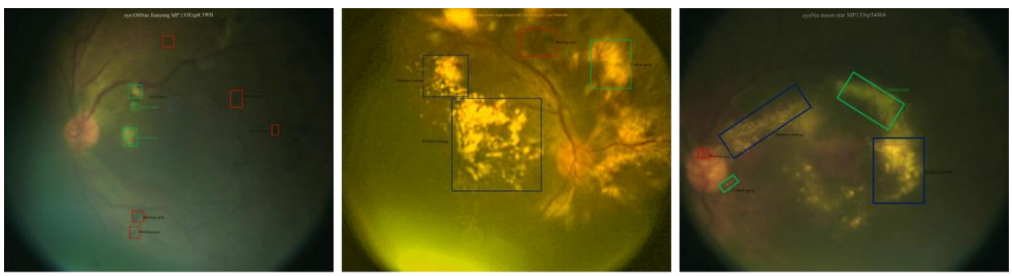
The model obtained 88.89% overall accuracy on the low-quality clinical test dataset. Tables 5 and 6 provide detailed prediction statistics for each DR grade, with specific testing details included in the Appendix. The model excelled in identifying DR0 and DR4 but faced challenges with DR3, as some DR3 images were misclassified as DR2. Figure 11 show cases the specific lesions in misclassified DR3 images.

Table 5. Data volume statistics for the clinical data test set

Clinical datasets	DR0	DR1	DR2	DR3	DR4	SUM
Number of images	18	0	9	6	3	36

Table 6. Model Prediction Statistics

Clinical datasets	DR0	DR1	DR2	DR3	DR4	SUM
Number of images	18	0	13	2	3	36



(a) DR3 judged to be DR2 (b) DR3 judged to be DR2 (c) DR3 judged to be DR2

Figure 11. Model misclassification image (DR3) lesion presentation.

An in-depth analysis of the experimental results suggests that the misclassifications may be due to the quality discrepancies between the clinical test dataset and the training dataset. The lower resolution, suboptimal lighting conditions, and extraneous characters at the top of some images likely affected the model's recognition accuracy. Additionally, the small size of the clinical test dataset may have constrained the model's generalization ability, affecting the statistical significance of the results. Therefore, future research should expand the clinical test dataset

and improve the model's robustness when handling low-quality images to enhance its practical value and reliability.

DISCUSSION

This study introduces a remote DR screening system called REACH, designed to overcome challenges in traditional screening methods. The system integrates a portable fundus image acquisition device, a deep learning inference module, and an information management system. This REACH system significantly reduces equipment costs and dependence on specialized technicians. It also optimizes the screening process, allowing patients to complete their initial screening in community or home settings without multiple hospital visits. The system is particularly suited for regions with limited healthcare resources, improving the coverage and efficiency of DR screening. The REACH system demonstrated high accuracy and strong practicality in public datasets and small-scale clinical trials, showing significant potential for widespread future adoption. The system has already been deployed at Zhejiang Provincial People's Hospital for online testing, with plans to extend its use within the primary healthcare system. Future developments aim to integrate primary screening with secondary and tertiary hospitals, creating a platform that leverages early intervention for blindness prevention and treatment.

The REACH system is designed on a multi-layered, modular architecture to ensure scalability and adaptability. The system comprises five core subsystems: the FDAS, DES, AIPS, DSS, SIMS, and IQAS. FDAS enables high-quality, rapid image acquisition through optimized optical design. DES incorporates a certificate-free signcryption algorithm at the data collection end, unifying the signing and encryption processes. This design is well-suited for resource-constrained regions. AIPS uses neural networks such as EfficientNetV2-s, DenseNet-121, and ResNet-50 to classify fundus images and improve classification accuracy by incorporating relationally weighted labels. SIMS manages all screening-related data, ensuring data security, traceability, and multi-role user access control. The system's tech stack is built on a Vue front-end, Spring Cloud microservices, and MySQL database, providing flexibility for cross-platform compatibility and the ability to meet diverse hardware deployment requirements.

Despite its promising capabilities, the REACH system still faces several limitations:

- Dataset limitations: The current model relies on publicly available datasets and small-scale clinical data, leading to insufficient dataset diversity, particularly regarding race, geography, and device variations. This lack of diversity limits the model's generalization ability, which may lead to performance degradation in more complex real-world scenarios.
- Accuracy and generalization: Although the system demonstrates high accuracy in experimental environments, its generalization ability when handling low-quality or complex images requires further improvement, particularly in extreme cases.
- Challenges in real-world deployment: The system may encounter infrastructure challenges during widespread implementation, especially in remote areas with unreliable network connections and limited acceptance of new technologies, which may impact its effectiveness.
- Limited clinical data: The current clinical tests are small in scale. Although the results indicate the system's practicality, robustness, and stability in larger clinical environments, further validation is still needed.

Future research should prioritize expanding the dataset, including data from different regions and ethnicities, to enhance the model's generalization ability. Further investigation is needed to improve the model's robustness in processing low-quality images and handling complex real-world scenarios. A key objective for the future is to facilitate the widespread use of the system widespread use in remote areas. This scenario involves extending the system to more communities and families in collaboration with local healthcare institutions to increase early DR screening coverage.

CONCLUSION

In conclusion, the REACH system offers a series of key innovations and improvements compared to existing DR screening methodologies. It establishes a comprehensive and integrated healthcare ecosystem by encompassing the entire workflow—ranging from image acquisition to AI-driven diagnoses—and proves particularly effective

in primary care contexts. The portability of the fundus image acquisition device, paired with mobile applications, facilitates screenings not only in community settings but also in home environments, thereby optimizing traditional processes and enabling large-scale deployments. By supporting multiple user roles—including doctors, patients, and administrators—the system achieves enhanced applicability and usability for a diverse range of stakeholders. Moreover, its adoption of a front-end and back-end separation strategy, along with a microservices architecture, promotes flexibility and supports cross-platform operation across various hardware and operating systems. This design effectively bolsters system scalability and maintainability.

Real-time data management and an integrated feedback mechanism further strengthen the screening workflow by enabling instant diagnostic information to be relayed to clinicians, thus enhancing both efficiency and accuracy. The Certificate-Free Signature Encryption Algorithm, which merges signing and encryption into a single operation, ensures the confidentiality and integrity of sensitive information, effectively protecting patient privacy. Meanwhile, deep learning model optimization through the introduction of self-attention mechanisms and relationally weighted labels significantly refines DR classification performance. Finally, small-scale clinical testing has verified the system's robust predictive capacity even when dealing with low-quality images, laying a solid foundation for broader implementation and real-world clinical applications.

ACKNOWLEDGEMENTS

This study was funded by the Key Research and Development Projects of Zhejiang Science and Technology Plan (No. 2021C03103).

REFRENCES

[1] World Health Organization. (2016). Global report on diabetes. Geneva, Switzerland: World Health Organization.

[2] Sabanayagam, C., Banu, R., Chee, M. L., Lee, R., Wang, Y. X., & et al. (2019). Incidence and progression of diabetic retinopathy: A systematic review. The Lancet Diabetes & Endocrinology, 7(2), 140–149.

[3] Chinese Diabetes Society, National Diabetes Prevention and Management Office. (2022). National grassroots diabetes prevention and management manual (2022). Chinese Journal of Internal Medicine, 61(07), 717-748. https://doi.org/10.3760/cma.j.cn112138-20220509-00350

[4] Abou Taha, A., Dinesen, S., & Vergmann, A. S. (2024). Present and future screening programs for diabetic retinopathy: A narrative review. International Journal of Retina and Vitreous, 10(14). https://doi.org/10.1186/s40942-024-00534-8

[5] Padilla Conde, T., Robinson, L., & Vora, P. (2023). Effectiveness of telemedicine diabetic retinopathy screening in the USA: A protocol for systematic review and meta-analysis. Systematic Reviews, 12(48). https://doi.org/10.1186/s13643-023-02201-9

[6] Zang, P., Gao, L., Hormel, T. T., Wang, J., You, Q., Hwang, T. S., & Jia, Y. (2021). DcardNet: Diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography. IEEE Transactions on Biomedical Engineering, 68(6), 1859–1870.

[7] Usher, D., Dumskyj, M., Himaga, M., Williamson, T. H., Nussey, S., & Boyce, J. (2004). Automated detection of diabetic retinopathy in digital retinal images: A tool for diabetic retinopathy screening. Diabetic Medicine, 21(1), 84–90.

[8] Sinthanayothin, C., Kongbunkiat, V., Phoojaruenchanachai, S., & Singalavanija, A. (2003). Automated screening system for diabetic retinopathy. In Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis (pp. 915–920). IEEE.

[9] Torrent-Solans, T., Duarte, L., Monteiro, R., Almeida, E., Bernardes, R., & Cunha-Vaz, J. (2004). Red-dots counting on digitalized fundus images of mild nonproliferative retinopathy in diabetes type 2. Investigative Ophthalmology & Visual Science, 45(13), 2985.

[10] Ribeiro, L., Oliveira, C. M., Neves, C., Ramos, J. D., Ferreira, H., et al. (2015). Screening for diabetic retinopathy in the central region of Portugal: Added value of automated disease/no disease grading. Ophthalmologica, 233(2), 96–103.

[11] Karnowski, T. P., Abramoff, M. D., Niemeijer, M., Brea, D., Russell, S. R., et al. (2012). Automated image analysis and the application of diagnostic algorithms in an ocular telehealth network. In K. Yogesan, L. Goldschmidt, & J. Cuadros (Eds.), Digital teleretinal screening (pp. 43–57). Springer, Berlin Heidelberg.

[12] Giancardo, L., Meriaudeau, F., Karnowski, T. P., Li, Y., Garg, S., et al. (2010). Quality assessment of retinal fundus images using elliptical local vessel density. In D. Campolo (Ed.), New developments in biomedical engineering (Chapter 11, pp. 201–223). INTECH.

[13] Giancardo, L., Meriaudeau, F., Karnowski, T. P., Li, Y., Garg, S., et al. (2010). Microaneurysms detection with the radon cliff operator in retinal fundus images. In Medical Imaging 2010: Image Processing, Proceedings of SPIE (Vol. 7623).

[14] Abràmoff, M. D., Niemeijer, M., & Russell, S. R. (2010). Automated detection of diabetic retinopathy: Barriers to translation into clinical practice. Expert Review of Medical Devices, 7(2), 287–296.

[15] Niemeijer, M., Abràmoff, M. D., & van Ginneken, B. (2006). Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. Medical Image Analysis, 10(6), 888–898.

[16] Barriga, E. S., Murray, V., Thompson, H., Abramoff, M. D., & Soliz, P. (2010). Automatic system for diabetic retinopathy screening based on AM-FM, partial least squares, and support vector machines. In Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 1349–1352).

[17] Murray, V., Barriga, E. S., Murillo, S., Pattichis, M. S., Soliz, P., et al. (2012). Recent multi-scale AM-FM methods in emerging applications in medical imaging. EURASIP Journal on Advances in Signal Processing, 2012(23).

[18] Agurto, C., Barriga, E. S., Murray, V., Nemeth, S., Crammer, R., et al. (2011). Automatic detection of diabetic retinopathy and age-related macular degeneration in digital fundus images. Investigative Ophthalmology & Visual Science, 52(8), 5862–5871.

[19] Dupas, B. (2010). Evaluation of automated fundus photograph analysis algorithms for detecting microaneurysms, haemorrhages, and exudates, and of a computer-assisted diagnostic system for grading diabetic retinopathy. Diabetes & Metabolism, 36(2), 213–220.

[20] Decencière, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J.-C., et al. (2013). TeleOphta: Machine learning and image processing methods for teleophthalmology. IRBM, 34(2), 196–203.

[21] Quellec, G., Lamard, M., Cazuguel, G., Cochener, B., & Roux, C. (2013). Multimedia data mining for automatic diabetic retinopathy screening. In Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 7144–7147).

[22] Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. Proceedings of the 38th International Conference on Machine Learning.

[23] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4700–4708.

[24] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778.

[25] Jin, W., Meng, X., & Wu, Y. C. (2022). A review of deep learning applications in image classification. Modern Information Technology, 6(16), 29-31+35. https://doi.org/10.19850/j.cnki.2096-4706.2022.16.008

[26] Han, Z., Yang, B., Deng, S., Li, Z., & Tong, Z. (2023). Category weighted network and relation weighted label for diabetic retinopathy screening. Computers in Biology and Medicine, 152, 106408. https://doi.org/10.1016/j.compbiomed.2022.106408

[27] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324. https://doi.org/10.1109/5.726791

[28] Sklearn Pre-processing: Encoding categorical features. Retrieved from https://scikit-learn.org/stable/modules/preprocessing.html#encoding-categorical-features

[29] Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., et al. (2019). Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Information Sciences, 501, 511–522.

[30] Kaggle Aptos Blindness Detection Competition. (2019). Retrieved from https://www.kaggle.com/competitions/aptos2019-blindness-detection (Accessed April 4, 2022).